제5차 전자기록관리 연구포럼

지능형 전자기록관리를 위한 기반 기술 이해

2019.10.10

전북대학교
오 효 정

전북대학교

목차

1. 지능형 전자기록 서비스 선진 사례 조사
2. 지능형 정보처리 기반 기술
3. 전자기록관리를 위한 지능형 정보처리 기술 적용

1. 지능형 전자기록 서비스 선진 사례 조사

1.1 지능형 기술 유형 분석
1.2 사례 소개
1.3 시청각기록 특화 기술 분석
1.4 시사점

1.1. 지능형 기술 유형 분석

기록정보서비스 (외부고객: 일반 이용자)	검색 고도화	• 텍소노미/폭소노미(Taxonomy/Folksonomy) • 자동 태깅(Auto-Tagging)
	추천 서비스	• 빅데이터 분석(Bigdata Analytics) – 데이터 마이닝, 텍스트 마이닝 • 챗봇(Chat Bot)
	콘텐츠 제공	• 링크드 오픈 데이터(Linked Open Data), 온톨로지(Ontology) • 스마트 디바이스(Smart Device) – 사물인터넷(IoT) 디바이스, 웨어러블(Wearable) 디바이스, VR/AR 디바이스, 로봇(Robot) 디바이스
기록관리 업무지원 (내부고객: 업무 담당자)	기록정리	• 클러스터링(Clustering) • 기계 학습(Machine Learning)/딥러닝(Deep Learning)
	기술(Description)	• 문서요약(Text Summarization) • 키워드/메타데이터 추출
	이관 및 보존	• 클라우드 컴퓨팅(Cloud Computing) • 블록체인(Block Chain)

자연어 처리(NLP: Natural Language Processing)
빅데이터 처리

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

1. 지능형 전자기록 서비스 선진 사례

1.3 시청각기록 특화 – 기술 유형 분석

음성 관련 기술

이미지 관련 기술

동영상 관련 기술

기반 기술

자동 분류

객체 추출

비디오 요약

파일 속성 추출

TTS(Text To Speech)

SST(Speech To Text)

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

전북대학교

1. 지능형 전자기록 서비스 선진 사례

1.3 시청각기록 특화 – 선진 사례

음성 관련 기술	Audio Ontology	• 특정구간에 내용을 설명하는 메타데이터를 붙여 검색이 가능하도록 함 • 총 632개의 이벤트 클래스의 확장 온톨로지가 구축되어 있음
이미지 관련 기술	MMIO	• Multi-Modal Incompleteness Ontology Model • 이미지와 그와 관련된 텍스트 정보를 기반으로 정보 결합 및 이미지 추출을 위한 온톨로지 생성
	Adam 프로젝트 Microsoft	• 음성인식과 이미지 분류를 위한 엄청난 양의 데이터를 처리하는 딥 러닝 시스템 • 멀티미디어 [illegible] 대규모 시스템으로의 확장성, 심층 기계학습을 통한 빠르고 정확한 인식 가능
	iMARS 프로젝트 IBM	• 그 나의 상황 혹은 객체가 아니라 전체의 특징점 분포를 분석하여 객체 및 상황을 이해하는 기술 • 기본 인식 대상은 객체, 사람, 장면, 활동의 4가지 범주를 대상
동영상 관련 기술	Video Intelligence Google	• 동영상 내의 특정 사물 [illegible] 자동으로 찾고 그 객체와 장면 설명에 대한 표제어를 제시 • 객체와 [illegible] 매칭 확률까지 제공하며, 현재 API 제공 중
	X-Brain 프로젝트 Google	• 유튜브에서 무작위로 크롤링된 이미지 데이터를 대상으로 심층 학습을 수행 • 21,841개 카테고리 1,400만개 이미지를 인식한 결과 15.8%의 인식률(기존 연구 대비 약 70% 향상)
	DARPA 영상인식 프로젝트	• 대규모 감시 비디오 데이터에서 관심 있는 콘텐츠를 [illegible] 검색하고 특정 행위를 자동 인지하는 분석 도구 • 객체보다는 행동 인식에 초점을 맞추었으며 현재 23가지 행동을 인지함
	V live NAVER	• AutoCut기능(영상에서 특정 인물이 등장하는 구간을 모아서 보여주는 서비스)과 AutoCam기능(영상 내 인물들 중 특정 인물만을 따라다니며 보여주는 서비스)가 제공되고 있음 • 다만 특정 연예인에 대해서만 한정적으로 제공되며, 이미 편집된 영상을 서비스함

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

1. 지능형 전자기록 서비스 선진 사례

1.4 시사점

선진 사례 기술 활용 현황 비교

• 국내외 기록관, 도서관, 박물관의 지능화 선진 기술 활용 분석 결과를 종합적으로 비교한 것으로 활용 현황을 단계별로 표시

구분	세부 기술 내용	국외*			국내*		
		기록관	도서관	박물관	기록관	도서관	박물관
업무 지원 영역	• 클러스터링 • 기계 학습/딥러닝	△	X	△	△	X	X
	• 문서 요약 • 키워드/메타데이터추출	△	X	X	X	X	X
	• 클라우드 컴퓨팅 • 블록체인	○	△	X	○	○	X
정보 서비스 영역	• 텍소노미/폭소노미 • 자동 태깅	○	○	△	X	△	△
	• 시청각 검색 지원	△	X	X	△	X	X
	• 빅데이터 분석	X	△	○	X	○	X
	• 챗봇	△	△	○	△	△	X
	• LOD, 온톨로지	X	○	△	△	○	X
	• 스마트 디바이스	X	○	○	△	△	△

* X: 선진 기술 미도입 단계 / △: 선진 기술 도입 초기단계 / ○: 선진 기술 발전단계

업무지원 영역

- 국내외 모두 기록관 관련 선진기술 도입 활발
- 신기술을 통한 기록 정리 및 기술 관련 업무 지원 적극적
- 도서관 및 박물관의 경우 일부 신기술을 도입하고 있으나 활용도는 기록관에 비해 낮은 편

정보서비스 영역

- 여가제로 운영되는 기록관은 도서관에 비해 정보서비스 영역에서 지능화 선진 기술 도입 수준이 낮음
- 정보서비스 영역이 특화된 도서관 분야에서는 서비스 관련 선진 기술 도입이 활발하고, 특히 LOD 기반의 서비스 제공 수준이 높은 편

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

전북대학교

1. 지능형 전자기록 서비스 선진 사례

1.4 시사점

시청각기록 선진 사례 기술 현황 분석

신기술사례

- Audio Ontology
- iMARS 프로젝트 (IBM)
- Video Intelligence (Google)
- MMIO
- X-Beam 프로젝트 (Google)
- Adam 프로젝트 (Microsoft)
- Naver V live
- DARPA 영상인식 프로젝트

선진사례

- NAVCC
- AP Archives
- SBS OASYS
- 농촌진흥청
- 국방과학연구소

- IT분야 대기업들이 기계학습을 기반으로 한 동영상 관련 신기술 개발을 시도
- 특히, Audio Ontology와 Video Intelligence는 시범 서비스와 API를 통해 기술을 공개하고 있어 접근 및 사용이 가능
- 최근 기술은 이미지와 동영상에서 인물, 사물, 건물, 행위 등에 대한 정보를 추출하는 객체 추출에 초점을 맞추고 있으며, 이를 통해 자동 태깅(labeling)를 시도하고 shot(or scene)을 구분
- 현재 국내 기록관 현황은 기술 도입 초기 수준

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

2. 지능형 정보처리 기반 기술

2.1 인공지능
2.2 자연언어이해
2.3 자연언어이해 기반 응용 기술
2.4 미디어 분석 기술

2.1 AI: 인공지능 기술 개발 사례

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

2.1 AI: 인공지능 기술 개발 사례 – AlphaGo 동작 원리 (Deep Learning)

바둑 인공지능? 정의하자면:

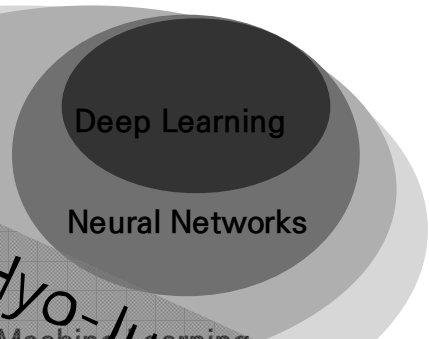
s'

- **Deep Learning**
  - 병렬 컴퓨팅 환경하에서
  - 경우의 수를 줄이고
  - 추상화하여
  - 최적화된 해답을 찾는 기계학습 기법

http://www.slideshare.net/ShaneSeungwhanMoon/ss-59226902

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

2.1 AI: 인공지능 – ML vs. DL



기계학습

- 분류에 효과적인 자질을 선정(feature engineering)하는 것이 중요
- 연구 초점:자질 확장 또는 효과적인 가중치 계산 방법 제안
- 문제점: 자질을 확장하기 위한 많은 시간과 노력이 필요

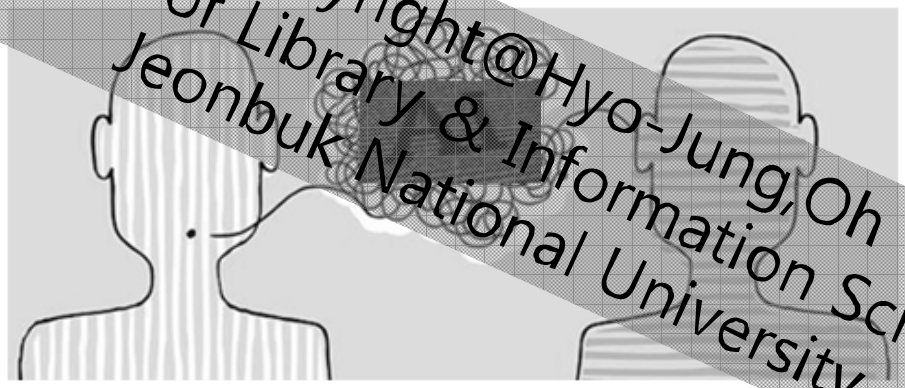
딥러닝

- 자질 선정에 대한 고민 해결
- 연구초점: 효과적인 분류를 위한 다양한 학습 모델 활용 및 자질 표현(word embedding) 방법
- 문제점: 대용량의 말뭉치, 성능향상의 명확한 원인 판단의 어려움

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

❑ 기계학습 기반 자연어처리(NLP: Natural Language Processing)

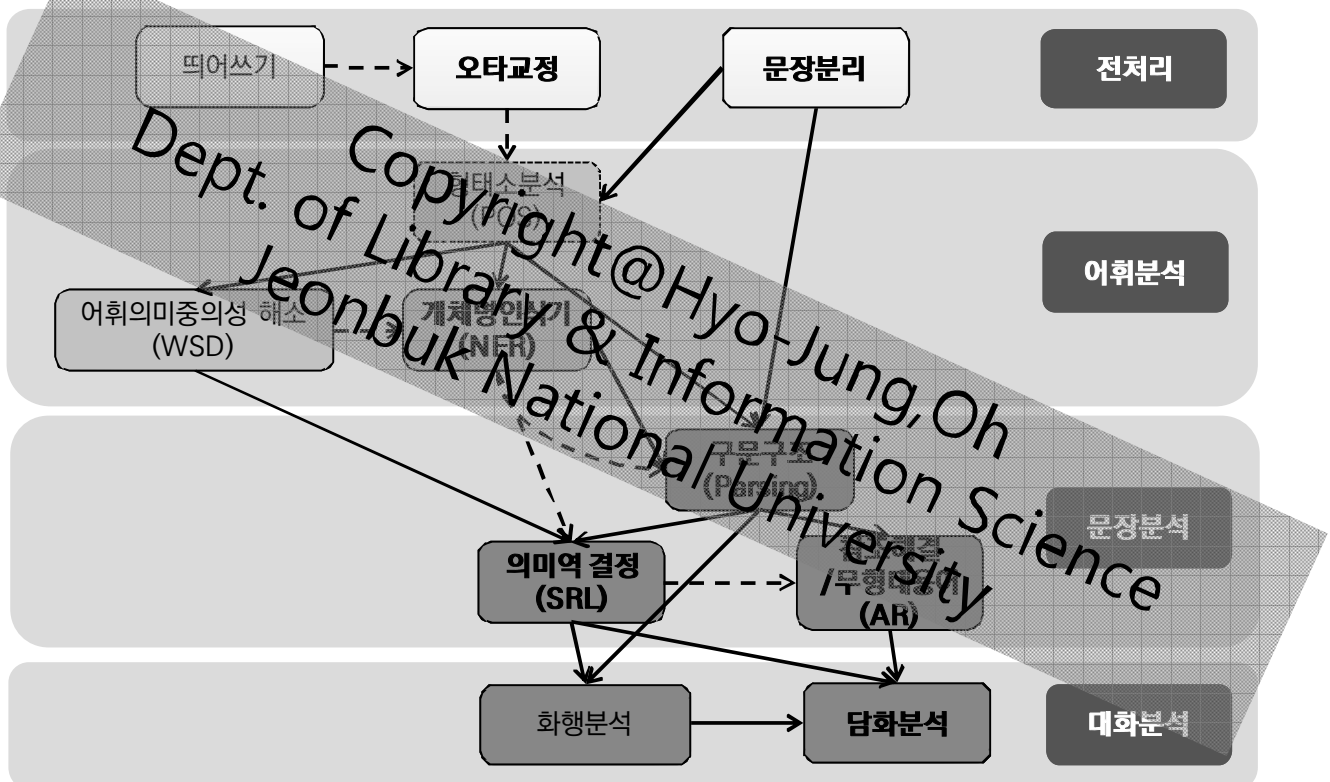
- 자연언어이해(NLU: Natural Language Understanding)
- 자연언어생성(NLG: Natural Language Generation)



Language computer language

Understanding Generation

Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

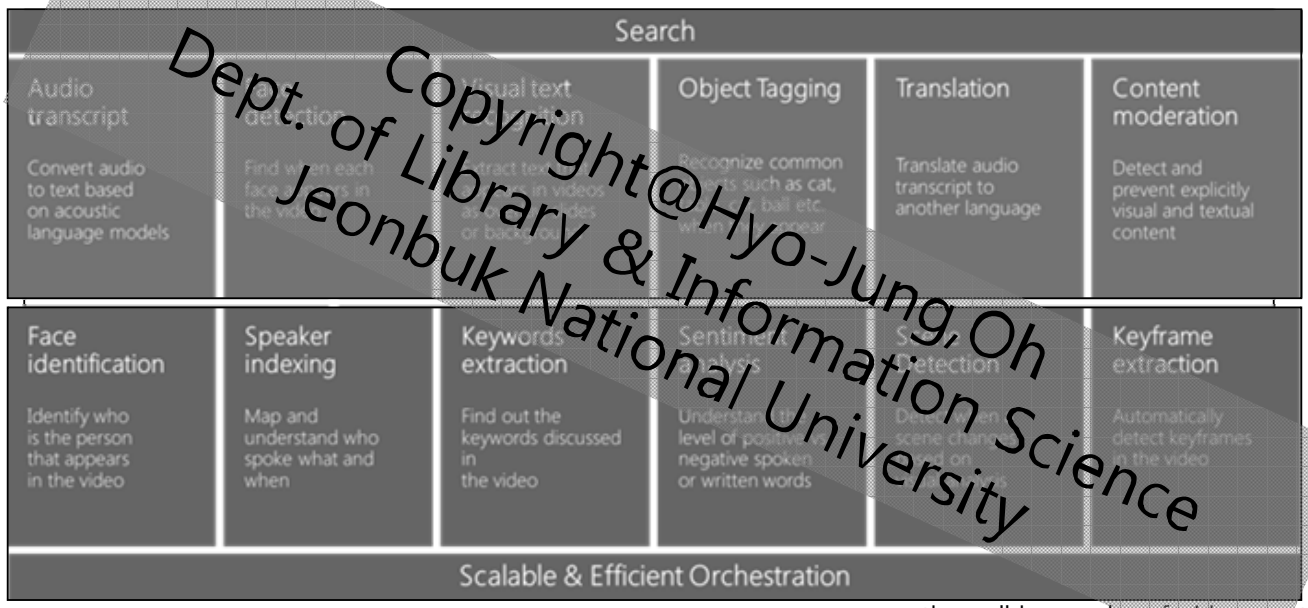


Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

→ 선행필수기술 - > 선행선택기술

2.4 미디어분석 기술 – Image & Video AI

❑ Video AI is the key to solving these challenges..



Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

…they all have a lot of video content!

❑ Video Indexer | Bundling Features Together

[출처] Azure Media Services & Video AI: Building next generation cloud video applications, *Microsoft*

3. 전자기록관리를 위한 지능형 정보처리 기술 적용

3.1 필요성
3.2 전자기록관리 업무 프로세스
3.3 프로세스별 지능화 요구사항 수렴
3.4 지능형 정보처리 기술 적용

3. 전자기록관리를 위한 지능형 정보처리 기술 적용

3.1 필요성

- 정부 업무환경이 변화함에 따라 **전자기록 수집대상을 확대**하고 **지능형 기록서비스 도입**, 정부 기록관리 기관의 역할을 재정립하고, 기록관리 법·제도를 재편하는 것이 시급하다는 의견이 논의되고 있음

- 특히, ICBM(IoT, Cloud, BigData, Mobile) 등 **4차 산업혁명의 핵심 기술들이 정부 업무에 도입되어 활용되고 있는 바**, 기록관리 분야에도 대량·대용량의 전자기록을 효율적으로 처리하기 위한 능동적인 전자기록관리 기반 마련이 시급함

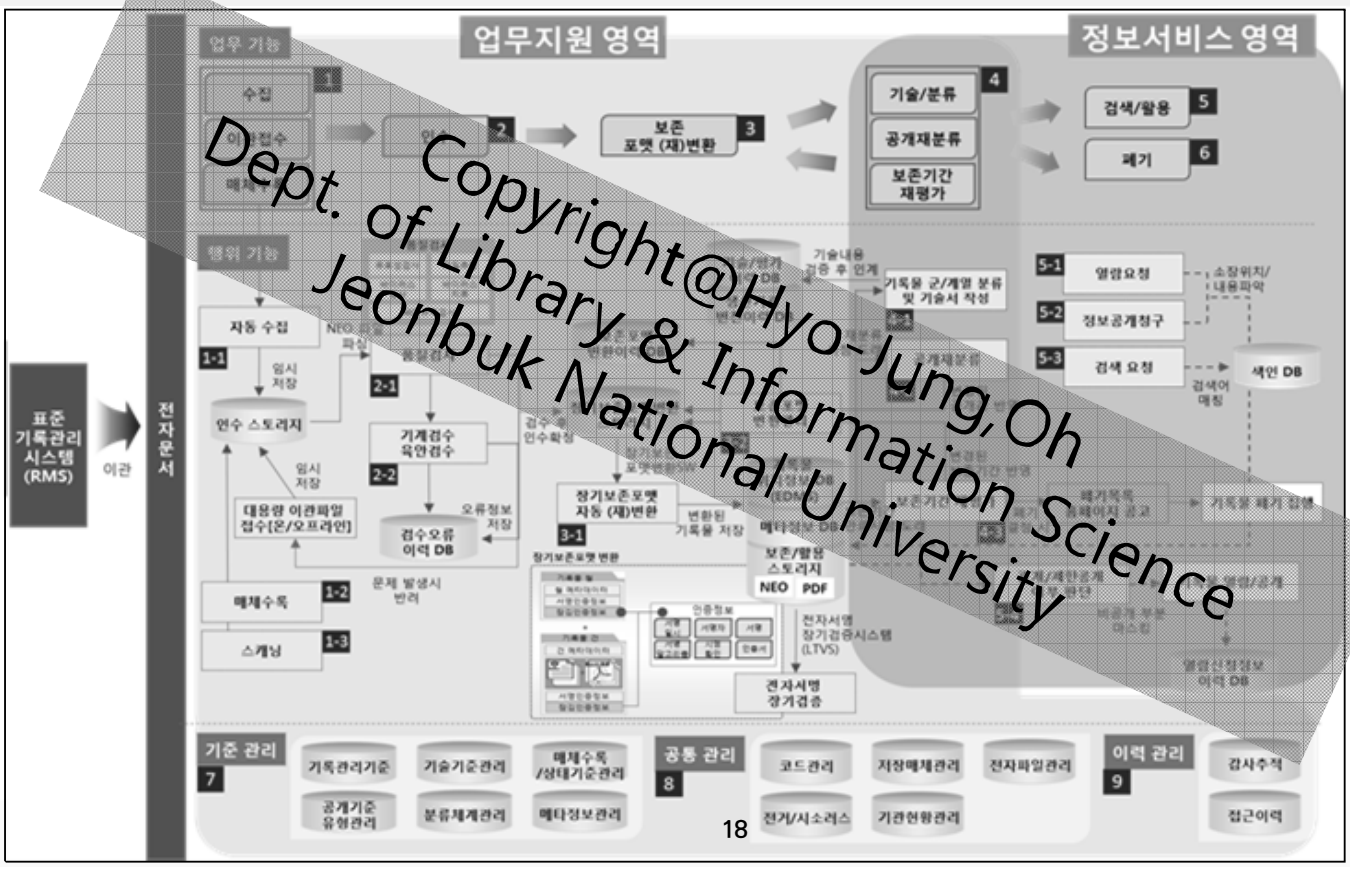
전자기록 생산 환경의 급격한 변화에 따른 관리대상 전자기록물의 급격한 증가와 유형의 다양화	현행 기록관리 업무처리 대부분이 수작업에 의존 -> 업무처리에 장시간이 소요되며, 미처리 기록물이 누적됨
대량·대용량 전자기록물의 메타데이터 생성 및 이관이 원활하지 않음	정리, 기술, 폐기 등의 기록관리 업무에 장시간이 소요
현재 전자기록관리체계는 표준전자문서(공문서)로 대표되는 전자기록물의 생산과 보존에 최적화	디지털 영상, 음성, 데이터세트, 웹기록 등 급증하는 신유형 전자기록물 수용 및 관리하기에 한계가 있음

기록관리에 인공지능 등 최신기술을 접목하여 업무를 개선하고,
자동화 기술 기반 대량·대용량 전자기록물의 효율적인 관리를 위한 기술 개발이 필요함

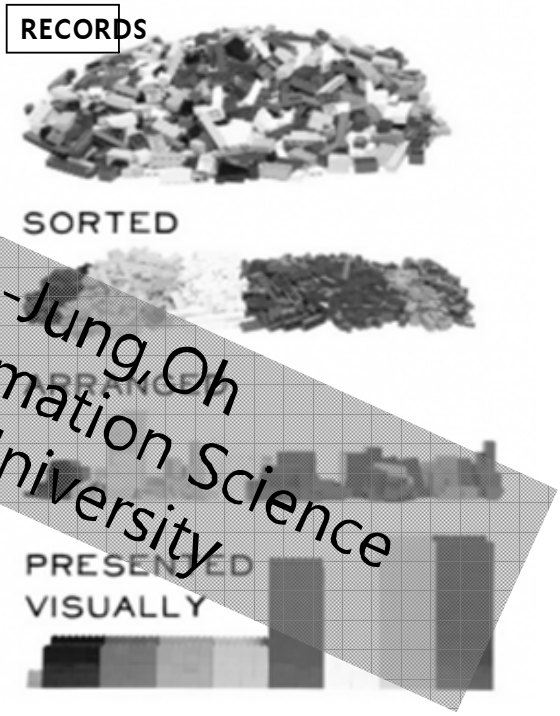
Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University

3. 전자기록관리를 위한 지능형 정보처리 기술 적용

3.2 전자기록관리 업무 프로세스 상세 분석



Copyright@Hyo-Jung,Oh
Dept. of Library & Information Science
Jeonbuk National University



Copyright@Hyo-Jung Oh
Dept. of Library & Information Science
Jeonbuk National University

Thank You
Q & A

Hyo-Jung Oh
(ohj@jbnu.ac.kr)